Chapter 8 Part II: Hypothesis Test Design

Significance vs. Power of Test

In a hypothesis test H_0 & H_A are determined by experiment results and α is determined by field of study / journal conventions.

→ The one thing you can control is $n = \#$samples

[Note: For most estimators $\hat{\Theta}$, the standard error $\sigma_{\hat{\Theta}}$ will shrink as n grows – the same result will have lower p-value (be more significant) if n is larger]

- If n is too small, test statistic may not reach α.
- If n is too big, waste time and money on unnecessarily large experiment.

The "significance" & "power" of a hypothesis test are measures of the probability of the test "failing" or "succeeding."

⇒ In many cases you can use them to pick the "best" value for n

There are two different ways that you could draw the wrong conclusion from a hypothesis test:

	Reject H_0	Fail to Reject H_0
H_0 True	Type I Error	OK
H_0 False	OK	Type II Error

Def: If H_0 should have been True but we (unluckily) got p-value $< \alpha$ (so we reject H_0) this is a Type I Error.

If H_0 should have been False but we (unluckily) got p-value $> \alpha$ (so we fail to reject) this is a Type II Error.

[Note: Type I Errors are worse than Type II Errors

- Type I Error ⟺ Reject a True H_0

  In this case you publish a False article.

- Type II Error ⟺ Fail to Reject a False H_0

  In this case you wasted time & money on experiment and missed interesting results.]

Thm: The significance of a test is P(Type I Error)

$$P(\text{Reject } H_0 \mid H_0 \text{ is True}) = \alpha$$

This is the same 'significance level' α from before ("Reject if p-value $< \alpha$")

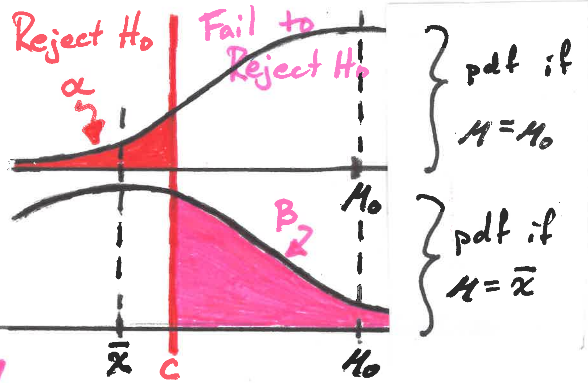
Def: $P(\text{Type II Error}) = \beta$

i.e. $P(\text{Fail to reject} \mid H_0 \text{ False}) = \beta$

The power of a test is $(1-\beta)$

$$\text{Power} = (1-\beta) = P(\text{Reject } H_0 \mid H_0 \text{ False})$$

Note: All other things being equal, making α smaller makes β bigger.



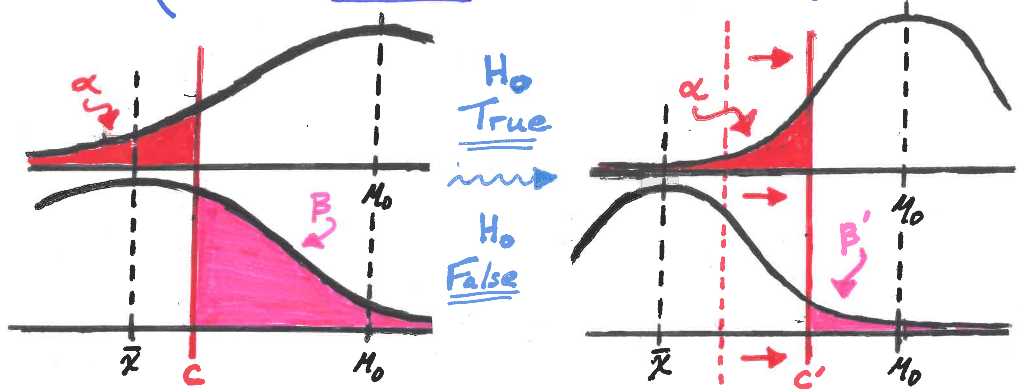
- $\alpha = P(\hat{\Theta} < c$ if H_0 True$)$
- $\beta = P(\hat{\Theta} > c$ if H_0 False$)$

(Reject if $\hat{\Theta} < c$
Fail to reject if $\hat{\Theta} > c$)

To compute β, we use $\theta = \hat{\theta}$, the value from our point estimate, (in this case $\hat{\theta} = \bar{x}$)

We want β to be small (and Power ≈ 1)

→ Make β smaller by increasing n (which decreases standard error $\sigma_{\bar{x}}$)



Making n bigger makes $\sigma_{\bar{x}}$ smaller, so pdf for "H_0 True" ($\mu = \mu_0$) shrinks faster. This moves the cutoff for rejecting H_0 closer to μ_0 (and farther from $\bar{x}$). The pdf for "H_0 False" also shrinks faster — so new β is much smaller.

We can actually compute β (and best n) for the different types of hypothesis test that we've done so far!

Given n ⇝ Compute β
Given β ⇝ Compute n

§8.2: "z-Test" for $\mu = E[X]$

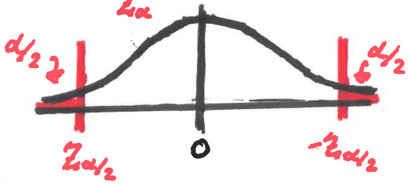
Recall: Test statistic is sample mean $\bar{X}$.

"Standardize" according to $H_0: \mu = \mu_0$

$$Z = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \qquad \left(\text{Write } z_{\bar{x}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right)$$

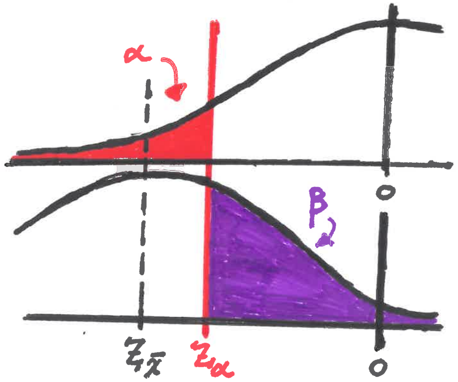
Cutoff for reject / fail to reject is

$\pm z_\alpha$ for one-tailed test



$\pm z_{\alpha/2}$ for two-tailed test

α & β for one-tailed test:



H_0 True: $\frac{\bar{X} - \mu_0}{s/\sqrt{n}} \approx \text{Normal}(0,1)$

$\alpha = P(Z < z_\alpha)$

H_0 False: $\frac{\bar{X} - \mu_0}{s/\sqrt{n}} \approx \text{Normal}(z_{\bar{x}}, 1)$

$$\beta = P(Z > z_\alpha - z_{\bar{x}})$$
$$= 1 - \text{pnorm}(z_\alpha - z_{\bar{x}})$$
$$= \text{pnorm}(z_{\bar{x}} - z_\alpha)$$

Example: Suppose sample 50 times & get $\bar{x} = 20$, $s = 6$. One-tailed test against $H_0: \mu = 22$ with $\alpha = .05$. What is β?

$$z_{\bar{x}} = \frac{20-22}{6/\sqrt{50}} \approx -2.357$$
$$z_\alpha = \text{qnorm}(.05) \approx -1.645$$

$$\beta = \text{pnorm}(z_{\bar{x}} - z_\alpha) \approx .238$$
$$\text{Power} = 1 - \beta \approx .762$$

Not Good!

Example: Same setup but sample 100 times.

$$z_{\bar{x}} = \frac{20-22}{6/\sqrt{100}} \approx -3.333$$
$$z_\alpha = \text{qnorm}(.05) \approx -1.645$$

$$\beta = \text{pnorm}(z_{\bar{x}} - z_\alpha) \approx .046$$
$$\text{Power} = 1 - \beta \approx .954$$

Good!

Given β we can solve for n:

$$\beta = P(Z < z_{\bar{x}} - z_\alpha)$$
$$z_\beta = z_{\bar{x}} - z_\alpha \qquad \text{Recall: } z_{\bar{x}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$
$$z_\alpha + z_\beta = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$
$$n = \left[\frac{s(z_\alpha + z_\beta)}{\bar{x} - \mu_0}\right]^2$$

Example: Suppose we expect $\bar{x} \approx 20$ and $s = 6$

Want to do one-tailed test against $H_0: \mu = 22$

with $\alpha = .01$ and $\beta = .02$.

What value of n is necessary?

$$z_\alpha = \text{qnorm}(.01) \approx -2.326$$

$$z_\beta = \text{qnorm}(.02) \approx -2.054$$

$$n = \left[\frac{s(z_\alpha + z_\beta)}{\bar{x} - \mu_0}\right] \approx 172.67$$

$\Rightarrow n = 173$ (n is an integer)

Summary: $z_{\bar{x}} = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

$$\beta = \text{pnorm}(z_{\bar{x}} - z_\alpha)$$

$$n = \left[\frac{s(z_\alpha + z_\beta)}{(\bar{x} - \mu_0)}\right]^2$$

For two-tailed test replace z_α by $z_{\alpha/2}$

---

§8.3 "t-Test" for $\mu = E[X]$

Same setup: Test statistic is $\bar{X}$.

$$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \qquad \left(t_{\bar{x}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right)$$

$$\beta = \text{pt}(t_{\bar{x}} - t_\alpha, n-1) \qquad (t_\alpha = \text{qt}(\alpha, n-1))$$

In this case, we cannot solve for n because $T \sim t(n-1)$ # degrees of freedom of T changes when n changes!

Most statistics programs can numerically find n

$\Rightarrow$ In R: power.t.test

§8.4 Population Proportion (test for p)

Setup: Test statistic is $X/n = \hat{p}$

$$\hat{p} \approx \text{Normal}\left(p, \sqrt{\hat{p}\hat{q}/n}\right) \qquad \hat{q} = (1-\hat{p})$$

$$n = \left[\frac{\sigma_{p_0} z_\alpha + \sigma_{\hat{p}} z_\beta}{p_0 - \hat{p}}\right]^2$$